

**ATS**OHIO
UNIVERSITYTEXAS A&M
KINGSVILLE**Ambient Fine Particulate
Research****PM_{2.5}**

Task 1.1: Data Inventory/Benchmarking for Database Applications

Benchmarking:

- CARB Data Management Project
- Supersites Integrated Relational Database for Supersites and Coincidental Data

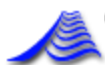
Inventory:

- UORVP Data
- SCAMP Data
- Pittsburgh Supersite Data
- NETL Site Data
- Other Potential Data Sources

**ATS**OHIO
UNIVERSITYTEXAS A&M
KINGSVILLE**Ambient Fine Particulate
Research****PM_{2.5}**

CARB Data Management Project

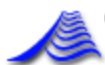
- The project has been underway for three years.
- To date the entire project, including data collection efforts, has cost a total of approximately forty-four million dollars.
- The data collection process lasted fourteen months.
- The data engine used is MS SQL Server.
- The program incorporates Quality Assurance Automation.

California Regional Particulate Air Quality Study
A Multi-Agency Collaboration

**ATS**OHIO
UNIVERSITYTEXAS A&M
KINGSVILLE**Ambient Fine Particulate
Research****PM_{2.5}**

CARB Data Management Project

- The program has integrated ad hoc query capabilities.
- The project is in the last/final phase of design.
- The database contains approximately one hundred million records.
- The program provides users with raw data.
- The interface still needs to be completed.



California Regional Particulate Air Quality Study
A Multi-Agency Collaboration



**ATS**OHIO
UNIVERSITYTEXAS A&M
KINGSVILLE

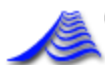
Ambient Fine Particulate Research

PM_{2.5}



Lessons Learned

- Need For Standards To Be Established
- Be Prepared To Assist Data Submitters With QA/QC And Formatting Operations
- Standard Input Formats Required For Any Large-scale Data Management Application
- Normalization Allows Different Kinds Of Data To Be Stored In Similar Table Structures



California Regional Particulate Air Quality Study
A Multi-Agency Collaboration



**ATS**OHIO
UNIVERSITYTEXAS A&M
KINGSVILLE**Ambient Fine Particulate
Research****PM_{2.5}**

Supersites Integrated Relational Database for Supersites & Coincidental Data

- University of Maryland-College Park and Environmental Protection Agency (EPA)
- Goal: To consolidate all Supersite data into centrally-managed database
- Effort was still in its embryonic stages



Ambient Fine Particulate Research

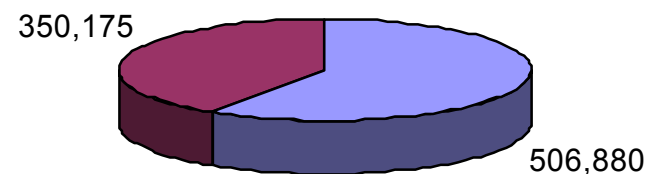
PM_{2.5}



UORVP Data

	Continuous Parameters	Discrete Parameters	Continuous Records	Discrete Records
Holbrook Primary	9	115	207,360	95,335
Athens Satellite	0	115	0	16,905
Lawrenceville Primary	13	115	299,520	220,225
Morgantown Satellite	0	115	0	17,710
TOTALS			506,880	350,175

Record Sources



Description	record size (Avg.)	No. of Records	Size
	Bytes (B)		Gigabytes (GB)
UORVP Combined	5000	857,055	4.29

[Summary Data](#)





Ambient Fine Particulate Research

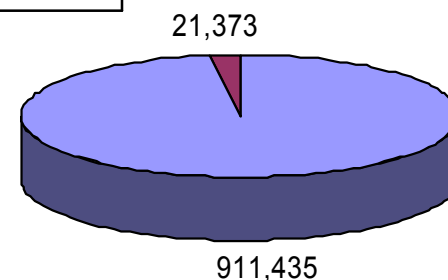
PM_{2.5}



SCAMP Data

	Continuous Parameters	Discrete Parameters	Continuous Records	Discrete Records
Steubenville Primary	33	34	911,435	10,215
TRSP Satellite	0	26	0	3,192
St Vincent College	0	26	0	3,190
Wheeling Jesuit University Satellite	0	26	0	3,297
Hopedale Satellite	0	26	0	1,479
TOTALS			911,435	21,373

Record Sources



Description	Record Size (Avg.)	No. of Records	Size
	Bytes (B)		Gigabytes (GB)
SCAMP Combined	5000	932,808	4.7

[Summary Data](#)





Ambient Fine Particulate Research

PM_{2.5}



Pittsburgh Supersite Data

Site Name	Number of Parameters Tracked		Start Date	End Date	Number of Records	100%
Pittsburgh Supersite	Continuous:	18	July 2001	October 2002	4,335,408	
	Discrete:	0			0	

Description	Record Size (Avg.)	No. of Records	Size
	Bytes (B)		Gigabytes (GB)
Pittsburgh Supersite	5000	4,335,408	21.68

NETL Site Data

Site Name	Number of Parameters Tracked ²		Start Date	End Date	Number of Records	100%
NETL Supersite	Continuous:	3	Jan 2001	Dec 2001	13,196	
	Discrete:	0			0	

Description	Record Size (Avg.)	No. of Records	Size
	Bytes (B)		Gigabytes (GB)
NETL	5000	13,196	0.07

[Summary Data](#)



**ATS**OHIO
UNIVERSITYTEXAS A&M
KINGSVILLE**Ambient Fine Particulate
Research****PM_{2.5}**

Other Potential Data Sources

- Pennsylvania Department of Environmental Protection (PADEP)
- Ohio Environmental Protection Agency (OHEPA)
- West Virginia Department of Environmental Protection (WVDEP)
- Allegheny County Health Department (AHCD)
- National Oceanic Atmospheric Administration (NOAA)

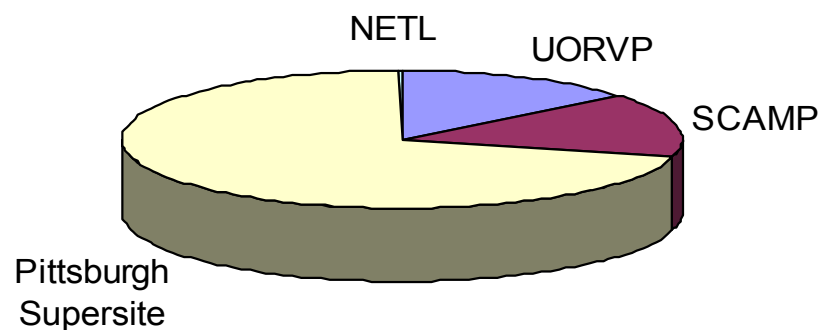




Inventory Summary

Program Name	Parameter Type	Number of Records ^{1,2}
UORVP	Continuous:	506,880
	Discrete:	350,175
SCAMP	Continuous:	911,435
	Discrete:	21,373
Pittsburgh Supersite	Continuous:	4,335,408
	Discrete:	0
NETL	Continuous:	13,196
	Discrete:	0
SRI	Continuous:	0
	Discrete:	0
TOTAL RECORDS	ALL	6,138,467

Data Sources



¹ Values represent estimates of observation data only, based on data submitted to date.

² Values represent submitted quantities only.